What is at Stake with High Stakes Testing? A Discussion of Issues and Research[1]

GREGORY J. MARCHANT, Department of Educational Psychology, Ball State University, Muncie, IN 47306

ABSTRACT. High stakes tests are defined as those tests that "carry serious consequences for students or educators." The consequences from standardized achievement tests range from grade retention for school children to rewards or punitive measures for schools and school districts. The nature of standardized achievement tests used in these situations poses validity problems for the decisions. Numerous unintended negative consequences for students, teachers, curriculum, and schools have been identified. Research has yet to establish clear benefits from these high stakes practices. Therefore, with little empirical support and financial and human costs high, a costs/benefits analysis suggests that the high stakes testing bandwagon, further fueled by No Child Left Behind, needs to be carefully evaluated before it continues to roll.

OHIO J SCI 104 (2):2–7, 2004

INTRODUCTION

For both advocates and opponents of the use of standardized tests in decisions regarding students, teachers, and educational policies, the answer to "what is at stake with high stakes testing?" is the same. The answer is, "everything." In an effort to implement accountability measures for districts, schools, teachers, and even individual students, testing originally designed to provide information regarding individual student achievement and ability for diagnostic/prescriptive teaching efforts is now being used as the measuring stick for evaluating the success of students, teachers, schools, districts, and even states. With important decisions resting on the results of certain test scores, it is important to know how well the scores reflect the quality of learning and education. It is also important to consider whether decisions based on these tests tend to reflect accurate interpretations and result in best practice. Even a potentially useful tool for education may be considered inappropriate if its use routinely results in harm to children.

This article began as a review of the current research to explore the results of high stakes testing; of particular interest was its affect on student learning. Surprisingly and unfortunately the impact of high stakes testing on student achievement has not been investigated. Therefore, this article reviews the research and concerns addressed in the literature regarding high stakes testing.

DEFINITION OF HIGH STAKES TESTING

A position statement issued by the American Educational Research Association issued in July of 2000 described high-stakes testing as follows:

> Many states and school districts mandate testing programs to gather data about student achievement over time and to hold schools and students accountable. Certain uses of achievement test results are termed "high stakes" if they carry serious consequences for students or educators. Schools may be judged according to the school-wide average scores for their students. High school-wide scores may bring public praise or financial rewards; low scores may bring public embarrassment or heavy sanctions. For individual students, high scores may bring a special diploma attesting to exceptional academic accomplishment; low scores may result in students being held back in grade or denied a high school diploma.

The statement then identified the 1999 Standards for Educational and Psychological Testing as guidelines for high-stakes testing efforts. The guidelines include protection against high-stakes decisions based on a single test, full disclosure of likely negative consequences of high-stakes testing programs, alignment of the test and the curriculum, opportunities for remediation for those who fail, appropriate attention to language differences and disabilities.

High-stakes tests are usually national or state-wide standardized achievement tests. If a test is "standardized" it has set rules for administration, such that everyone taking the test receives the same exact directions and has the same restrictions of time and resources. Achievement tests are usually for one specific grade level and designed to create a distribution of scores. Popular national standardized achievement tests are the Terra Nova and the Stanford-9. Many states have taken up the costly task of developing their own state achievement tests aligned with their state's standards. Some of these tests were developed in conjunction with national test makers and share items. The SAT is not an achievement test, but an aptitude test designed to predict college achievement; however, because of its influence on college admissions decisions, it is also considered a high-stakes test.

THE NATURE OF STANDARDIZED ACHIEVEMENT TESTS

Most standardized achievement tests are norm-referenced, in that how well an individual does on the test is based on a comparison to a large group of test takers. "Good" is relative to others at the same grade level. This is in contrast to a criterion-referenced test

[1]Manuscript received 5 December 2002 and in revised form 14 May 2003 (#02-29).

that defines how well one does on a test based on the meeting of criteria or mastering a standard. High stakes decisions tend to involve either relative comparisons or reaching a pre-defined cut-off point. However, almost always the decision as to where the cut-off point will be is informed by norm-referenced information, such as difficulty levels of items selected or even percentile rank of a score. Such that, if a cut-off score equates to the 40th percentile, the decision makers know that approximately 40% of the test-takers will not "pass" the test. Therefore, the setting of the cut-off score is very important on high-stakes tests that require passage. For example if a state like Ohio, that averages 140,000 students at each grade level, was to raise a cutoff score for a required achievement test by 5 percentiles, approximately 7,000 more children would *not* reach the cutoff at *each* grade level.

There are several problems inherent in standardized achievement tests as the basis for high stakes decisions (Popham 1999). Test designers, desiring a good distribution of scores to be able to differentiate students, do not want too many items that almost everybody gets right (or wrong). If just about everybody gets an item right it does not differentiate among students. Therefore, basic skills items that are important for everyone to master (and many do master them) are unlikely to show up on the test in large numbers. Therefore, some of the most important basic skills are not given much attention in tests. Due to limitations of time, the number of items measuring any particular skill or knowledge may be too few to provide a reliable measure of a specific skill. A strength or weakness may be determined by a few good guesses or a few skipped items. Time constraints and restriction in the range and nature of the items (usually multiple-choice responses) suggest that, although an achievement test can provide some information, as a one-time paper and pencil assessment it has serious limitations in measuring the variety and scope of classroom learning.

IMPACT ON STUDENTS

It is probably safe to say that the process of taking a large standardized achievement test does little to improve the knowledge or skills of students. Without feedback as to correct or incorrect responses, the exercise is one of demonstrating knowledge and skills rather than learning. The feedback the students receive from the test also has limited potential for improving their learning, skills, or knowledge. It is weeks, or more likely months, after the students fill in the last "bubble" with their #2 pencil that they or their teachers see the results. At that point the only meaning attached to the results can be reduced to how well they did, and in particular did they do well enough to avoid any negative consequences associated with not passing or reaching a particular cutoff score. If the students are old enough, the results can provide them with a means of comparing themselves with their peers. Then the results indicate whether they are "smarter" than their friends, or perhaps more importantly are they "dumber." Therefore, the impact that students feel from the testing process and resulting feedback is only indirect. It is the consequences and concern regarding those consequences that impact students, and those consequences are significant. The results from standardized tests may decide whether students pass to the next grade level or are retained, can establish whether students are eligible for certain special programs, may determine whether students graduate from high school, or may decide to which college students will be admitted. These are events that have a major influence on students' lives.

The practice of retaining students, thereby repeating a grade level, has been thoroughly studied, and the evidence is clear. Holding children back and simply having them repeat a grade level that they "failed" is bad policy with devastating consequences (for a recent policy statement from the National Association of School Psychologists see Anderson and others 2002). The effect of high stakes testing was evident in Baltimore last summer where a new promotion policy was based on passage of the Terra Nova national achievement test and state functional exams (Bowie 2002). More than a quarter of Baltimore's elementary and middle school students, over 20,000 students, were required to repeat a grade level in school after not meeting the requirements. Based on what we know about retention, many of these students, and students like them across the country, will feel the effects of high-stakes testing for the rest of their lives. An increase in dropout rates is considered a given as a result of high-stakes testing programs (Futrell and Rotberg 2002). The 300% increase in middle school dropout rates in five years in Boston has been attributed to high-stakes testing policies and rigid and indifferent responses to kids at-risk (Hayward 2002). The fact that minorities tend to do worse on standardized achievement tests means that they are more at-risk to experience the effects of testing policies. For example in Texas in less than a ten-year period, between 100,000 and 200,000 minority students would have stayed in school and received a diploma, if the minority passage rates were equal to non-minority students on the mandatory achievement test (Haney 2001).

> *During the summer that I moved back to Indiana from Ohio, a young girl stopped in my backyard to play on the swing set I had built. She exuberantly announced that, "today is the happiest day of my life!" I asked why, anticipating that she had been given a pony, pet, or some prized toy. Instead, she proclaimed that she had re-taken, and this time passed, the state achievement test, and would now be able to go on to the next grade level with her friends. At the time, I feared for the educational policy of the state I was to reside, now I fear for the educational policy of the country.*

It is because of what is at stake that students learn to value or fear standardized tests. Students come to devalue learning and schooling, and shift their emphasis to, "Is this going to be on the test?" (Paris 2000). Although young children tend to hold standardized achievement

testing in fairly high regard and not that different than regular classroom tests, negative perceptions and distinctions between regular tests and standardized tests increase by grade level, and are most pronounced among high school students and low achievers (Paris and others 2000; Wong and Paris 2000). These negative attitudes of resentment, anxiety, cynicism, and mistrust of testing were found to manifest themselves in test taking behaviors like loafing, cheating, and stress related behaviors. In 1984 (Hill), an estimated 10 million students in elementary and secondary schools performed below ability on tests because of anxiety. With the increased emphasis and importance on testing, that number is likely to be much higher today. Over a third of the Arizona and Texas teachers surveyed reported that as a reaction to high stakes tests their students experienced headaches, upset stomachs, irritability, increased aggression, and "freezing" during parts of the test (Morison 1992). Fewer teachers also reported students crying, truancy, vomiting, and refusal to take the test.

IMPACT ON TEACHERS AND CURRICULUM

An important question is, does high stakes testing change what and how teachers teach and, if so, does it change instruction for the better? It is hoped that high stakes achievement tests will encourage teachers to focus on meaningful achievement areas and improve those areas in their students. It is also believed by many that the cumulative scores of a teacher's students reflect the quality of the instruction and can provide a basis for accountability. Teachers' beliefs regarding high stakes testing appear relatively universal across the United States (Haladyna and others 1991; Heubert and Hauser 1999; Hoffman and others 2001; Smith 1991a); Urdan and Paris 1994). Teachers believe there are too many tests, the results are not useful to teachers and are misunderstood by parents and the public, and the tests are unfair to minorities and English as a Second Language students (Paris and Urdan 2000). The majority of teachers think that standardized achievement tests are not worth the money or instructional time that they cost (Urdan and Paris 1994).

> A review of the larger literature initially suggested that state-mandated testing both positively and negatively influenced teachers' beliefs and practice. Once my gaze focused on those works that could be identified as research, however, empirical support for the claim that state-mandated testing positively influences teachers' beliefs and practice seemed to vanish.
>
> (Cimbricz 2002, p 6)

Effect on Form and Content of Instruction

Those hoping that high-stakes testing would lead to a "back-to-basics" approach in terms of content taught and teaching approaches used might deem the research results as positive. Research repeatedly yields two findings, teachers tend to narrow the scope of their curriculum to that which is tested, and they tend to abandon more innovative teaching strategies, such as cooperative learning and creative projects, in favor of more traditional lecture and recitation (for example, Brown 1992, 1993; Romberg and others 1989). Because of the publication of test scores and the implications for the quality of teaching, teachers feel compelled to teach-to-the-test in hopes of improving their students' scores (Smith 1991a). Those areas not tested, often science, social studies, health, and even writing, are neglected in favor of reading and arithmetic skills that appear on the tests. High stakes testing also seems to encourage the use of instructional approaches and materials that resembles testing (Rottenberg and Smith 1990). Rituals of multiple-choice quizzes and test preparation take the place of "normal" instruction. Teachers exploring instructional practices informed by current views of learning and cognitive psychology that seek deeper understanding and critical thinking, may find those techniques and even those goals at odds with the drill and practice suggested by the broad superficial coverage typical of achievement tests.

> *When sixth grade scores dipped across the state of Indiana, middle schools responded (even though the statewide drop was likely attributable to the nature of that particular test). In my daughter's school, five minutes were taken from each class period to create a period for students who failed the test to get remediation. Unfortunately, the rest of the students were left with a half-hour less instruction to sit in a study hall.*

Effect on Test Preparation Approaches

Although some test preparation is to be expected, the amount of time devoted to such activities and the nature of the test preparation speaks to the stress created by high stakes testing. Teachers across the country spend varying amounts of time preparing their students for high stakes tests. Time that previously was devoted to learning skills and knowledge in an appropriate sequential fashion, gets lost in the process of cramming for the tests. In Arizona (Nolen and others 1992), elementary teachers reported spending more time than was required in test preparation activities (33%), some started preparation two months before the tests (28%), and a few gave daily test practice over two weeks before the tests (10%). In contrast some teachers (22%) reported giving no advance test preparation, leading to concerns that the variability in preparation could be confounding scores (Smith 1991b).

The Arizona study (Nolen and others 1992) found that along with practicing appropriate test preparation strategies, elementary teachers made a point of covering specific topics from the tests (66%), used commercial test-preparation packages (41%), taught vocabulary on the test (26%), used practice items from the previous year's tests (12%), and even taught items from the current year's tests (10%). In Texas it was determined that teachers from lower scoring schools reported a higher incidence of questionable test preparation approaches, as well as blatantly unethical practices during testing such as giving hints, pointing out mistakes, giving instruction,

and directly identifying correct answers (Hoffman and others 2001).

IMPACT ON SCHOOLS, DISTRICTS, AND STATES

An essential consideration for any test, as well as for any research study, is validity. For a test to be valid, it must actually measure that which it purports to measure. If a standardized achievement test is to be a valid measure of student learning, the quality of instruction of a teacher, or the effectiveness of the educational system of a school, district, or state, that test must match the curriculum being taught. If the state adopts a national standardized achievement test that is not aligned with the state's curriculum, the test is not a valid measure of educational quality in the state. In fact, less effective teachers or schools that stray from the state's standards to focus on test content could appear more proficient than competent teachers efficiently teaching the standards. A panel of 30 reviewers (2 principals, 18 teachers, and 10 parents) evaluated a nationally-marketed third grade achievement test (Bauer 2000). Although almost all of the items were deemed important for students to know, approximately 50% of the items were judged as inappropriate due to bias (primarily based on socio-economic level) or because the content was viewed as not part of standard third grade curriculum. This mismatch of achievement tests and curriculum is not a new concern. An important study in 1983 (Freeman and others) reviewed the content of five nationally standardized achievement tests in mathematics in grades 4 through 6 and the content of four widely-used textbooks for the same grade levels. In no case was even 50% of the test content adequately addressed in the textbooks and, for some tests, 80% of the items did not receive meaningful attention in the textbooks. A more expensive alternative to simply adopting a national achievement test is for the state to develop its own test designed to be a valid measure of student mastery of the state's standards. Although states developing their own tests work to align the tests with their standards, the match between what is taught and what is measured (especially on national achievement tests) is still a major concern.

Research studies strive to eliminate alternative hypotheses for the results of the study to be considered valid. Current outcome-based evaluation approaches look to the same rigors as research in attempting to test results. Therefore, if judgments are to be made concerning the relative quality of teachers, schools, districts, or states, it is important that the results are valid measures of quality instruction and effective policies. If significant student achievement differences can be attributed to something other than educational quality, then the validity of judgments concerning educational quality come into question. A major threat to validity in educational research is selection bias. Selection bias may occur when the samples being compared are not randomly assigned to different treatment groups, such that the sample in the treatment groups are qualitatively different from each other in ways that can impact the results. Then comparisons of the treatments (in this case instruction and educational policies) are confounded by the differing nature of the samples.

Children are not randomly assigned to states, school districts, schools, and often not even to teachers. There are qualitative differences in students that are not the result of instructional quality or educational policies. When students with like characteristics, known to be related to achievement, are over- or under-represented in samples, then those factors are difficult to ignore when making judgments concerning the groups' achievement. However, those suggesting the use of raw aggregated achievement test scores as a means of evaluating teachers, schools, districts, and states are doing just that; they are ignoring the very real and present differences that exist in groups of students that are outside of the control schools.

Every August the College Board publishes the average SAT scores for each state and the District of Columbia and, despite warnings regarding potential misinterpretation, every August newspapers across the country publish front-page stories indicating the rise or fall of their state in the ranks. However, over 90% of the difference in aggregated SAT scores among the states is due to characteristics that the education systems have no control over (such as parent education and income; Marchant and Paulson 2001). It is mostly the aggregated characteristics of the test-takers that is being compared across states, not the quality of their education system.

The effects of student characteristics are evident from SAT scores down to elementary school standardized achievement scores. In Ohio (Ohio Department of Education 2002), the passage rate on the state proficiency test is 10 to 39% lower for schools with 50% or more of their students receiving free-and-reduced lunches. Compared to whites, the passage rate for African-American students ranges from 18% lower in the 10^{th} grade to 40% lower in the 6^{th} grade. Schools with a higher percentage of poor and/or African-American students are likely to have significantly lower average scores regardless of the quality of the instruction in the schools. Although it is possible to statistically control for student differences due to demographic variables and other factors out of teachers' control, as has been done for years in Tennessee (Sanders and Horn 1994), many view these statistics as too complicated or as suspicious manipulations of the data.

Of course the easiest way for a state to change its level of success is to change the criteria for success. Instead of placing the bar high and working to reach that goal over a number of years, states fearing public ridicule and potential repercussions from the "No Child Left Behind" legislative mandates may revisit their definition of "proficient." For example (Hoff 2002a): Louisiana will consider students proficient if they score at the state's "basic" achievement level, Colorado will consider students proficient if they score at the state's category of partially proficient, and Connecticut will set its federal proficiency level lower than what is expected in the state's accountability system.

COSTS/BENEFITS ANALYSIS CONCLUSION

Never has there been a greater need for cost-benefit and cost-effectiveness analyses (Hummel-Rossi and Ashdown 2002), and nowhere is there a greater need than in determining whether the benefits of high stakes testing justify the costs. These costs come not only at a financial level but also at a personal level for the children, and at a professional level for educators. At the financial level the costs are high. States that had planned to implement some tests are dropping or postponing those plans because of economic problems and because of the "No Child Left Behind" mandate of testing every child in grades 3-8 in reading and math by 2006. The Oregon state department of education is scrapping its writing tests for 3rd, 5th, and 8th graders, its 5th and 8th grade science assessments, and the hand-scored extended response portions of its 5th and 8th grade math exams (Hoff 2002b). The move will save $4.5 million. In Missouri, individual school districts are deciding to pick up the $5.30 per student science and social studies testing cost, a move that will save the state $7.1 million. In Maryland, middle schools may choose not to administer the state's 8th grade test this year as long as they do not receive Title I money. Massachusetts is postponing giving its history and social studies tests.

Lynn Corno (2000) used the "Trojan Horse" metaphor to describe the high stakes testing movement. Like the Trojan horse, high stakes testing was welcomed into our school doors without knowing what harm was hidden inside. "Solid empirical data on the consequences of high stakes testing is overdue because this horse rolled into over half of the fifty states sometime ago (p 125)." With the "No Child Left Behind" federal mandate for testing, states are likely to be increasing the amount of testing they do. There are three points to consider in evaluating the role of standardized achievement tests in our schools:

1. There is little evidence that teachers' evaluations of students' learning are seriously flawed (for example, overall, high school grade point average is as good a predictor of freshman year success as the SAT, Bridgeman and others 2000).

2. There is no evidence that an extensive standardized testing program improves instruction, or more importantly student learning.

3. There are considerable negatives associated with high stakes testing in terms of potential damage to teachers and students, in the development of flawed policies, and in the financial burdens that divert time and money from instruction. These negatives demand the re-evaluation of increased testing and suggest the need to consider limitations on the weigh and consequences of these tests.

Whether high stakes testing is a "Trojan Horse" or just another bandwagon rolling through our educational system, there are enough concerns to put on the brakes, at least until the real impact on student learning can be assessed.

LITERATURE CITED

Anderson GE, Whipple AD, Jimerson SR. 2002. Grade retention: achievement and mental health outcomes. Bethesda: Nat Assoc Sch Psych. <http://www.nasponline.org/pdf/graderetention.pdf>. Accessed 1 Nov 2002.

Bauer SC. 2000. Should achievement tests be used to judge school quality? Ed Policy Analysis Arch 8 (46). <http://epaa.asu.edu/epaa/v8n46.html>. Accessed 26 Sep 2001.

Bowie L. 2002 (Aug 28). 20,000 children to repeat a grade. Baltimore Sun. <http://www.sunspot.net/news/education/bal-te.md.schools28aug28.story?coll=bal-home-headlines>. Accessed 30 Aug 2002.

Bridgeman B, McCamley-Jenkins L, Ervin N. 2000. Predictions of freshman grade-point average from the revised and recentered SATI: Reasoning test. Research Rept No. 2000-1 (ETS RR No. 00-1). New York: College Examination Board.

Brown DF. 1992. Altering curricula through state-mandated testing: Perceptions of teachers and principals. Paper at the annual meeting of the Amer Educ Resrch Assn. Apr 1992. San Francisco, CA.

Brown DF. 1993. The political influence of state-mandated testing reform through the eyes of principals and teachers. ERIC Document Reproduction Service No. ED 360 737.

Cimbricz C. 2002 (Jan). State-mandated testing and teachers' beliefs and practice. Educ Pol Analysis Arch 10 (2). <http://epaa.asu.edu/epaa/v10n2.html> Accessed 26 Feb 2002.

Corno L. 2000. Comments on Trojan horse papers. Issues in Ed 6:125-31.

Freeman DJ, Kuhs TM, Porter AC, Floden RE, Schmidt WH, Schwille JR. 1983. Do textbooks and tests define a national curriculum in elementary school mathematics? Elem Sch J 83:501-13.

Futrell MH, Rotberg IC. 2002 (Oct). Predictable casualties. Education Week 22 (5):34, 48. <http://www.edweek.com/ew/ewstory.cfm?slug=05futrell.h22> Accessed 10 Oct 2002.

Haladyna T, Nolen SB, Haas NS. 1991. Raising standardized achievement scores and the origins of test score pollution. Ed Researcher 20(5):2-7.

Haney W. 2001. Commentary response to Skrla et al. The illusion of educational equity in Texas: a commentary on 'accountability for equity.' Int J Leadership in Educ 4(3):267-75.

Hayward E. 2002 (Jun 19). Middle school dropout rate up. Boston Herald. p 3. <http://www.bostonherald.com/cgi-bin/www.bostonherald.com/search/search.bg> Accessed 21 Jun 2002.

Heubert JP, Hauser RM. 1999. High stakes: testing for tracking, promotion, and graduation. Washington (DC): National Academy Pr.

Hill KT. 1984. Debilitating motivation and testing: a major educational problem, possible solutions, and policy applications. In: Ames RE, Ames C, editors. Research on motivation 1. New York: Academic Pr. p 245-74.

Hoff DJ. 2002a (Oct 9). States revise the meaning of "proficient." Educ Wk 22(6):1, 24-5. <http://www.edweek.com/ewstory.cfm?slug=06tests.h22> Accessed 10 Oct 2002.

Hoff DJ. 2002b (Oct 9). Budget woes force states to scale back testing programs. Educ Wk 22(6):24. <http://www.edweek.com/ewstory.cfm?slug=06tests-sl.h22> Accessed 10 Oct 2002.

Hoffman JV, Assaf L, Paris SG. 2001. High stakes testing in reading and its effects on teachers, teaching, and students: today in Texas, tomorrow? Reading Tcher 54:482-92.

Hummel-Rossi B, Ashdown J. 2002. The state of cost-benefit and cost-effectiveness analyses in education. Review Educ Resrch 72:1-30.

Marchant GJ, Paulson SE. 2001. State comparisons of SAT scores: who's your test taker? NASSP Bull 85:62-74.

Morison P. 1992. Testing in American schools: issues for research and policy. Soc Policy Rep 6:1-24.

Nolen SB, Haladyna TM, Haas NS. 1992. Uses and abuses of achievement test scores. Educ Measure: Issues Practice 11(2):9-15.

Ohio Department of Education. 2002 (Mar). Ohio schools committed to success for all: 2002 annual report on educational progress in Ohio. Columbus (OH): Ohio Dept of Educ.

Paris SG. 2000. Trojan horse in the schoolyard: the hidden threats in high stakes testing. Issues Educ 6:1-16.

Paris SG, Roth JL, Turner JC. 2000. Developing disillusionment: students' perceptions of academic achievement tests. Issues Educ 6:17-45.

Paris SG, Urdan T. 2000. Policies and practices of high-stakes testing that influence teachers and schools. Issues Educ 6:83-107.

Popham WJ. 1999. Why standardized test scores don't measure educational quality. Educ Leadership 56(6):8-15.

Romberg TA, Zarinna EA, Williams SR. 1989. The influence of mandated testing on mathematics instruction: grade 8 teachers' perceptions. Madison (WI): Univ of Wisconsin, Cntr for Educ Resrch,

Sch of Educ, and Office of Educ Resrch and Improv of the US Dept of Educ.
Rottenberg C, Smith ML. 1990 (Apr). Unintended effects of external testing in elementary schools. Paper presented at the annual meeting of the Amer Educ Resrch Assn, Boston.
Sanders WL, Horn SP. 1994. The Tennessee Value-Added Assessment System (TVAAS) mixed-model methodology in educational assessment. J Personnel Eval Educ 8:299-311.
Smith ML. 1991a. Put to the test: the effects of external testing on teachers. Educ Resrchr 20(5):8-11.
Smith ML. 1991b. Meanings of test preparation. Amer Educ Resrch J, 28:521-42.
Urdan TC, Paris SG. 1994. Teachers' perceptions of standardized achievement tests. Educ Policy 8:137-56.
Wong CA, Paris SG. 2000. Students' beliefs about classroom tests and standardized tests. Issues Educ 6:47-66.